基于多源特征后融合的分层目标检测算法

盛 雷 卫志华 张鹏宇
(同济大学计算机科学与技术系 上海 201804)
(嵌入式系统与服务计算教育部重点实验室(同济大学) 上海 201804)

摘 要 目标检测是计算机视觉领域的热门研究课题,是视频内容分析的基础。文中提出了一种基于图像多源特征后融合的分层目标检测算法。在该算法中,使用多级决策的思想对目标检测任务进行粗细两个粒度的划分。在粗粒度层面,先使用HOG特征对图像进行分类,根据分类器的置信度分数,将测试图像分为正例、负例和不确定例。在细粒度层面,使用多种视觉特征以及多种核函数后融合的方法对不确定域中的图像做进一步分类。在同一数据集上设置了3组对比实验。实验结果表明,所提算法在各个评价指标上都有出色的表现,且在实际视频的目标检测中的效果优于Faster-RCNN。

关键词 计算机视觉,目标检测,多级决策,特征提取,后融合
中图法分类号 TP301.6 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.02.038

Multi-layer Object Detection Algorithm Based on Multi-source Feature Late Fusion

SHENG Lei WEI Zhi-hua ZHANG Peng-yu
(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)
(Key Laboratory of Embedded Systems and Service Computing, Tongji University, Shanghai 201804, China)

Abstract Object detection is a hot topic in computer vision and it is the foundation of video caption. This paper proposed a multi-layer object detection algorithm based on multi-source feature late fusion, and used ways of multi-level decisions to divide the object detection task into two granularities. At the coarse level, the HOG feature was used to classify the images. According to the confidence scores of the classifier, the test images were categorized into positive, negative and uncertain examples. At the fine level, this paper proposed a multi-source feature late fusion method to classify the examples which are in the uncertain field. This paper conducted several comparative experiments on the same data set. Experimental results demonstrate that the proposed algorithm can obtain excellent results in all evaluation metrics, and achieve a better detection result than Faster-RCNN.

Keywords Computer vision, Object detection, Multi-level decision, Feature extraction, Late fusion

1 引言

随着当代社会的快速发展,人们对安全防护的要求越来越高,监控视频已经被部署到城市的各个角落,因此视频图像已然成为当代社会中重要的大数据对象。在智能交通领域,行人和车辆作为视频图像内容的主要载体,对其进行检测与识别就变得尤为重要。一直以来,目标检测都是计算机视觉研究的热点问题,每年都有很多学者在目标检测方面发表最新的论文以及研究成果,尽管如此,当前的目标检测还是没有一套完备的通用算法用于实际的检测任务。因此,目标检测极具发展空间,仍然是计算机视觉领域中的重要研究课题。

传统的目标检测分为目标实例检测和目标类别检测[1]。目标实例检测是根据检测目标的表观特征判别性来进行区分的,无纹理特征或者纹理特征不明显的目标的视觉特征一般不能被可靠地提取到,因此,这类目标通常依据被检测目标的形状来确定。这种情况下,传统的检测方式是基于模板匹配的方法[2]及其扩展方法[3],由于是通过边缘提取算法[4]得到图像的特征,因此这种方法非常容易受到遮挡和背景的影响。为了弥补这种算法的不足,文献[5-6]提出了使用图像梯度方向作为目标检测的特征方法,但这种方法存在着一定的误报问题。之后,针对无纹理特征的目标检测方法越来越多,但是这些方法也都存在着各种各样的问题,文献[7]针对目标实例检测中存在的问题以及解决方法给出了详细的介绍。对于检测目标纹理特征比较可靠的情况,可以通过提取稳定的纹理

到稿日期:2018-07-13 返修日期:2018-10-19 本文受国家自然科学基金项目(61573259),公安部重大专项(20170004),国家重点研发计划项目(2017YFC0821300)资助。

盛 雷(1993-),男,硕士,主要研究方向为目标检测、图像处理;卫志华(1979-),女,副教授,博士生导师,CCF会员,主要研究方向为机器学习、数据挖掘、视频内容理解等,E-mail:zhihua_wei@tongji.edu.cn(通信作者);张鹏宇(1993-),男,硕士,主要研究方向为目标检测、目标跟踪等。

特征来进行检测。SIFT 特征[8]、特征描述子 PCA-SIFT[9]、SURF 特征[10]等的发展对纹理特征明显的目标检测发展起到了重要的作用。对目标类别的检测一直是众多研究者聚集的重要研究课题,文献[11]使用 HOG 特征结合 SVM 的方法,在行人检测方面取得了不错的结果。文献[12]则应用 Haar 特征在人脸检测方面取得了进展。但是,单一的特征分类器的效果总是差强人意,于是一种鲁棒性能较好的 AdaBoost 算法被应用到目标检测中[13]。近些年,随着深度学习的快速发展,Krizhevsky[14]提出了基于深度卷积神经网(DCNN)的图像分类算法,该算法不但提高了图像分类的准确率,而且也极大提升了目标检测的准确率。文献[15]提出将 DCNN 框架作为目标检测中的回归器预测输入;文献[16]则使用 DCNN 对图像中的目标进行回归检测,并给出了每个目标的置信度分数;文献[17]提出了一种新的 DCNN 框架,将目标检测识别任务集成起来,取得了不错的效果。之后,Girshick 等[18]提出了 RCNN 网络模型,其大幅提高了目标检测的效果,在 PASCAL VOC 2007 上的准确率在 66%～67%之间,但存在测试速度慢、训练时间长和训练所需的空间大等问题。Girshick[19]于 2015 年又提出了 Fast RCNN 模型,针对 RCNN 的问题给出了较好的解决方案,与 RCNN 相比,在准确率基本相同的情况下,该方案所需的训练时间和测试时间大幅缩短。之后,在 Fast-RCNN 的基础上,Faster-RCNN 模型[20]又被提出,该方法进一步缩短了测试时间,且该模型的框架更加精简。随之,基于 Faster-RCNN 的 Mask-RCNN[21]又被提出,该网络框架主要是在 Faster-RCNN 目标检测的基础上实现了图像语义分割的任务,大幅度提高了图像分割的精度。

无论是传统的目标检测算法还是近年来发展迅速的神经网络模型,都推动了目标检测的进步。但是,传统的目标检测算法的精确率不高,且存在着较多的误判。神经网络算法虽然大幅度提升了检测效果,却有着更长的训练时间且对计算机的硬件有着更高的要求。本文在传统的目标检测算法的基础上,提出了一种基于多源图像特征后融合的分层目标检测算法。

本文的主要贡献如下:

(1)使用分层分阶段决策的思想,对图像进行粗细两个粒度层的分类,以提高分类的正确率;

(2)通过提取多种图像特征来构造分类器,实验证明,在使用单个特征的情况下,HOG 特征的表现最优;

(3)提出了一种多源特征后融合的细粒度目标检测算法,实验证明,多种图像特征的使用可以解决单个特征表征能力不足的问题,多种核函数的混合使用可以解决单个分类器学习能力不足的问题。

2 基于特征后融合的目标检测算法

2.1 基本框架

使用 HOG-SVM 的方法,并结合重新学习(retraining)策略以及 Selective Search[22]算法来实现基本的目标检测,并将其作为后续实验的基本框架。

算法的基本流程如下:在分类器学习方面,首先将训练集分为正例和负例,对于训练集中的每一张图片,提取出图像的 HOG 特征,并将之合并成一个特征行向量,再根据该图像所属类别对得到的特征打上相应的标签,得到标签向量。将所有训练集图片特征向量和标签分别组合成特征矩阵和特征向量后作为分类器的输入,用于分类器的训练。在分类器的构造上,使用线性核函数的 SVM。训练得到的分类器不可避免地会有误报和漏报的问题,为增强分类器的学习能力,使用 retraining 的方法,将分类器中漏报和误报的图片提取出来,给定误报的图片修正标签－1,给定漏报的图片修正标签＋1,然后再将这些图片加入到之前的训练集中进行重新学习,得到一个新的分类器。在分类器训练完成后,生成候选框。利用传统的滑窗法获取候选框的时间复杂度较高,因此在本文框架中使用 Selective Search 算法生成较少的高质量的候选框,最后在这些候选框中使用分类器得到候选框的类别和位置信息。算法流程图如图 1 所示。

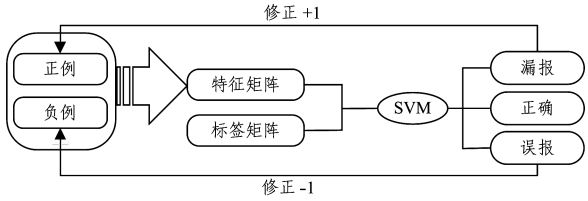


图 1 HOG＋SVM 算法

Fig. 1 HOG＋SVM algorithm

2.2 基于多源特征后融合的分级算法

本文的目标检测算法采用了多层分类和图像特征后融合的思想,如图 2 所示。第一层粗粒度的分类器采用基于 HOG 特征的单一分类器方法,第二层细粒度的分类器采用基于图像特征后融合的方法。对于输入的图片数据,经过第一层粗粒度分类器的决策得到该图像目标的置信度分数,将置信度分数划分为正域、负域和不确定域。对于置信度分数位于正域和负域内的数据,给出明确的判断;而对于置信度分数位于不确定域之内的数据,采用延迟决策的思想将该图片数据送入第二层细粒度的分类器进行决策后得到最终的分类结果。

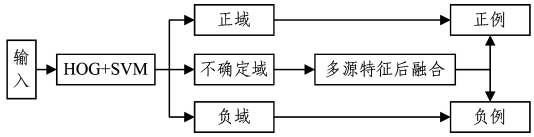


图 2 算法的整体框架

Fig. 2 Overall framework of algorithm

(1)多层图像分类算法

本文在进行目标检测的实验过程中发现,检测的结果中对于置信度分数较高和较低的情况,分类器往往都会给出正确的分类结果;而对于置信度分数处于中间水平的图片目标,往往存在着较多的误检和漏检。针对这种情况,本文使用 HOG 特征结合 SVM 的方法,对图片数据先进行分类,对于置信度分数位于中间水平的图片,将其标记为难例后输入到不确定域中。因此,根据置信度分数,本文将检测的结果划分为正域、不确定域和负域,每一部分对应着不同的决策规则。

定义 1 引入一对阈值 α 和 β 来定义正域、负域和不确定域中的图片目标对象,设 $0\leqslant\beta<\alpha\leqslant1$,则正域、不确定域和负域可定义为:

$POS_{(\alpha,\beta)}(X)=\{x\in U|P(X|[x])\geqslant\alpha\}$

$BND_{(\alpha,\beta)}(X)=\{x\in U|\beta<P(X|[x])<\alpha\}$

$NEG_{(\alpha,\beta)}(X)=\{x\in U|P(X|[x])\leqslant\beta\}$

对应正域、边界域和负域的定义,生成每一种情况中对应分类的图片要做的决策。对于每种情况生成不同的决策,每种决策的定义如下:

$P:\{x\in U|P(X|[x])\geqslant\alpha\}$

$B:\{x\in U|\beta<P(X|[x])<\alpha\}$

$N:\{x\in U|P(X|[x])\leqslant\beta\}$

规则 P 表示当置信度分数大于或等于 α 时,将该图片划分到正域中,即判定为正例,输出结果;规则 B 表示当置信度分数位于 α 和 β 之间时,将该图片划分到不确定域中,执行不确定域决策,即将图片输入到第二层分类器着重进行检测;规则 N 表示当置信度分数小于 β 时,将该图片划分到负域中,执行负域决策,即将图片数据判定为负例。

(2)多源特征后融合的图像分类算法

多源特征后融合的算法的流程如图 3 所示。该算法主要对上层图像分类算法中的不确定域中的数据进行检测。该部分图片数据中要检测的目标一般存在遮挡或者是目标较小的问题,因此较难检测。考虑到 SVM 分类器的分类效果是与其内核函数相关的,不同的特征与不同的核函数组合会产生不同的分类效果,本文采用多种特征结合线性核函数与单一特征结合不同核函数的方法,来对该部分图片数据着重进行检测。

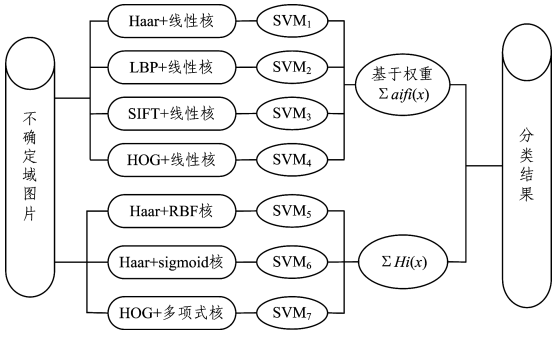


图 3　多源特征后融合算法的框架

Fig. 3　Framework of multi-source feature late fusion object detection algorithm

该算法的具体步骤如算法 1 所示。

算法 1　图像特征后融合算法

输入:不确定域中的图片

输出:若 result 为正数,则输出为正例,否则输出为负例

Step1　分别训练 Haar 特征、SIFT 特征、HOG 特征和 LBP 特征,并分别结合线性核函数得到分类器 $f_i(x)$;

Step2　计算每种分类器在测试集上的错误率 ϵ_i;

Step3　根据 ϵ_i 计算每个分类器的权重 α_i:$\alpha_i=\frac{1}{2}\ln(\frac{1-\epsilon_i}{\epsilon_i})$;

Step4　每个分类器得到预测结果 $f_i(x)$,将结果累加得到 $result_1$:$result_1=\sum_{i=1}^{4}f_i(x)*\alpha_i$;

Step5　分别训练 HOG 特征结合多项式核函数、RBF 核函数和 Sigmoid 核函数的分类器 $h_i(x)$;

Step6　每个分类器的预测结果为 $h_i(x)$,将结果进行累加得到 $result_2$:$result_2=\sum_{i=1}^{3}h_i(x)$

Step7　对两部分的结果进行累加,得到最终的结果:$reuslt=result_1+result_2$。

3　实验

3.1　数据集

本次实验所使用的训练图片为 COCO 数据集,该数据集是微软团队获取的一个可用来进行目标识别、图像分割的数据集,目前有很多目标检测的工作都是基于 COCO 数据集开展的。本次使用的测试集视频是在同济大学嘉定校区内拍摄的一段基于车载摄像头的路况视频。在该视频中,有步行、骑行、多人并排行走、遮挡等行人目标检测的复杂情形,因此将该视频作为测试集具有一定的挑战性。该视频总时长为 1 分 59 秒,视频帧的分辨率为 1920 * 1080,总帧数为 3570。由于行人姿态的多样性,如走路、骑车等多种形态,我们在实验中着重针对图像中的行人进行了检测。同时,还有一些额外的遮挡会增加检测难度,如并排行走、走路撑伞等情况。因此,行人检测一直是目标检测中的一个难点。本文中的算法针对视频中的行人着重进行了检测。

3.2　评价指标

对于图像特征分类器的分类效果,我们采用正确分类的正例(TP)和正确分类的负例(TN)来计算检测效果的正确率;同时,结合该分类器模型的准确率、召回率、F 值以及分类时间来对分类器进行综合评价。首先定义混淆矩阵,如表 1 所列。

表 1　混淆矩阵

Table 1　Confusion matrix

	预测类别 1	预测类别 0
实际类别 1	TP	FN
实际类别 0	FP	TN

混淆矩阵中,FP 表示实际图片中没有要检测的目标但是该分类器却检测到的结果;FN 表示实际图片中包含要检测的目标但是分类器没有正确检测到的结果。根据混淆矩阵,可以计算得到该分类器的正确率、准确率和召回率。计算公式如下:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}\tag{1}$$

$$Precision=\frac{TP}{TP+FP}\tag{2}$$

$$Recall=\frac{TP}{TP+FN}\tag{3}$$

根据式(1)－式(3)的计算结果可以得到 F 值:

$$F=\frac{2*Precision*Recall}{Precision+Recall}\tag{4}$$

对于本文中的目标检测实验结果,由于所采用的测试集为视频,因此采用的评价指标为标定数目、漏报数和误报数。其定义如下:

标定数目:实验结果中认为是目标的数量。

漏报数:未检测到的目标数量。

误报数:标定数目中是错误目标的数量。

3.3　实验设置

3.3.1　特征比较实验

为了探究单一图像特征对图像目标检测的表示能力,并为后续的集成分类器做准备,对相同的训练集提取 Haar,

LBP,SIFT 和 HOG 特征,并在同一个测试集上进行测试。此次实验的训练集包含正例(包含目标)200 张,负例(不包含目标)400 张;测试集包含正例 100 张,负例 200 张。分别计算每种特征分类器在测试集上的正确率、精确率、召回率和 F 值。

3.3.2 多源特征后融合的比较实验

采用与上述特征相同的训练集和测试集,计算在测试集上的正确率、准确率、召回率和 F 值。

3.3.3 目标检测对比实验

对比实验 1 为基于 HOG 特征的单分类器的基本模型,该方法属于传统的目标检测基本算法。对比实验 2 为基于 AdaBoost 的多特征多分类器的集成算法,该方法是对传统检测效果进行加强的经典算法之一。对比实验 3 为 Faster-RCNN 模型,该深度学习模型是对 RCNN 和 Fast-RCNN 模型的进一步改进,其在目标检测上的实用性已被证明,同时它也是现在目标检测的最主流方法。该模型原本只针对特定的 20 类目标进行检测。为了更好地进行对比,采用同一个训练集对原 Faster-RCNN 模型进行重新训练,得到了只检测单一目标的网络模型。虽然目前已经更新到 Mask-RCNN,但是 Faster-RCNN 的检测效果仍然十分惊人,与之对比足以说明问题。上述 3 个对比实验中,本文所采取的值为 $\alpha=0.7$,$\beta=0.3$(α 和 β 的值是在同一个数据集上经过多次实验得到的最佳组合结果)。对于其他的数据集,可以适当地增加 α 的值或降低 β 的值。

3.4 实验结果与分析

3.4.1 特征比较的实验结果

结合表 2 和图 4 可以看出,LBP 特征和 Haar 特征对图像目标的表示能力稍逊于 HOG 特征,二者表现在伯仲之间,而 SIFT 特征在本次实验中的效果最差。本次实验中对图像目标检测表现最好的是 HOG 特征,尤其体现在正确率、精确率和 F 值 3 个指标上,其召回率稍逊于 LBP 特征。因此,综合比较各种图像特征之后,使用 HOG 特征作为第一层粗粒度的分类器是合理的。

表 2 特征效果的比较

Table 2 Comparison of feature effect

特征	*Accuracy*	*Precision*	*Recall*	*F*
Haar	0.874	0.887	0.710	0.789
LBP	0.870	0.801	0.810	0.805
SIFT	0.787	0.687	0.660	0.673
HOG	0.907	0.950	0.760	0.844

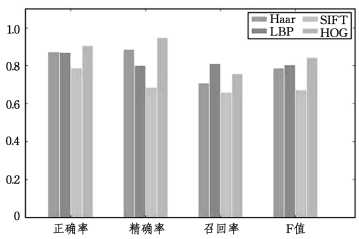


图 4 特征分类效果的比较

Fig. 4 Comparison of feature classification effect

3.4.2 多源特征后融合的实验结果

从图 5 中可以看出,本文提出的图像特征后融合的方式在正确率、精确率、召回率和 F 值 4 个指标上都较大幅度地高于 HOG 特征,从而证明了本文方法是有效的。

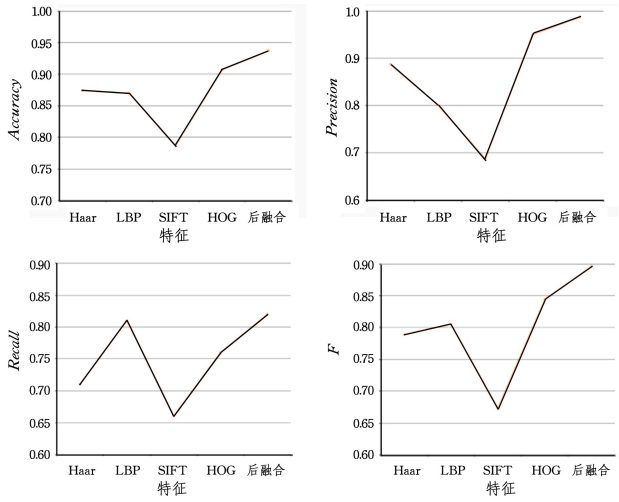


图 5 特征后融合结果比较

Fig. 5 Comparison of feature late fusion effect

3.4.3 目标检测的实验结果

基于 HOG 特征的单分类器算法的检测结果如图 6 所示。



图 6 实验 1 的结果

Fig. 6 Results of experiment 1

从图 6 中可以看出,该算法可以较好地检测出视频中行人目标的位置(图(c)和图(d)),但是存在着少许漏报(图(b)),同时存在着比较严重的误报(图(a),图(e)—图(h)),对树干、石柱等这类轮廓比较像人的干扰物并不能很好地区分。

基于 AdaBoost 的多特征多分类器的集成算法的检测结果如图 7 所示。Faster-RCNN 的检测结果如图 8 所示。



(a)　(b)　(c)　(d)　(e)　(f)　(g)　(h)

图 7　实验 2 的结果

Fig. 7　Results of experiment 2



(a)　(b)　(c)　(d)　(e)　(f)　(g)　(h)

图 8　实验 3 的结果

Fig. 8　Results of experiment 3

从图 7 中可以看出，该算法对于实验 1 中存在的误报问题有了较好的改善(图(e)和图(f))，能够正确地识别出相应的行人和车辆，但是误报问题依旧存在。另外，该算法并没有很好地将视频中的行人车辆目标检测出来(图(a)－图(b)，图(d)，图(h))，存在着较严重的漏报。

从图 8 可以看出，Faster-RCNN 可以较准确地识别出视频中各种形态的行人目标(图(a)－图(b)，图(f)－图(h))；几乎不存在误报问题，漏报数目也较少，但对于遮挡和目标较小的情况的检测效果并不太好(图(d)－图(e))。

本文算法的实验结果如图 9 所示。



(a)　(b)　(c)　(d)　(e)　(f)　(g)　(h)

图 9　本文算法的结果

Fig. 9　Experimental results of proposed algorithm

从图 9 中可以看出，本文算法模型能很好地区分视频中的行人和似人干扰物，误报大大减少。对于视频中一些较难识别的行人，如远处较小的人(图(d)－图(f))、骑摩托车的人(图(c))、并排行走的人(图(f)－图(g))等，本文算法模型都能很好地进行识别，漏报率较低。

对上面 4 个实验进行总结分析，统计出每种算法的准确标定目标数、漏报数和误报数，如表 3 所列。

表 3　结果比较

Table 3　Results comparison

算法	准确数	漏报数	误报数
HOG＋SVM	23	5	22
AdaBoost	14	14	5
Faster-RCNN	21	7	0
本文算法	26	2	2

从表 3 中可以看出，单分类器单特征加上重新学习的方

案可以识别出视频中各种形态的目标,但是这种方案的误报比较严重,几乎接近准确标定数目,误报率为 40%,漏报率为 17.9%,因此结果并不理想。基于 AdaBoost 的改进算法可以较好地减少误报数,误报率为 26.3%,但是存在着比较严重的漏报问题,漏报率为 50%,即使是一些比较明显的行人目标也没有被标定出来,因此结果亦不理想。Faster-RCNN 的识别精确率比较高,误报率为 0,但对遮挡和目标较小的情况存在少量的漏报,漏报率为 25%,总体效果很好。而本文提出的算法模型由于综合了多种特征,并且采用三支决策的分类方法,对视频中出现的行人目标都可以较好地检测到,且漏报率只有 7.1%,是参与对比算法中最低的;误报率为7.1%,优于对比实验 1 和实验 2。实验结果表明,本文算法模型对于视频中的行人目标检测是有效的。

结束语 本文提出了一种基于多源图像特征后融合的分层目标检测算法,通过对图片数据使用 HOG 特征先进行初级的分类;然后再结合多种特征和多种核函数的 SVM 对图片数据进行细粒度上的分类;最后在车载视频上进行了相应的实验,并做了 3 组对比实验来验证本文所提算法的有效性。实验表明,本文算法在该测试数据集上取得了令人满意的效果。本文所提出的多源特征后融合的分层目标检测算法本质上是一种集成学习的方式,通过组合多种不同的模型来解决单一分类器效果不佳的问题。本文使用多种特征可以丰富对图像的表达,集成多种核函数是为了提升分类器的学习能力和分类能力。集成学习的方法是一种具有普适性的思想,因此绝不会只对本文中所做的实验有效。对于其他目标的检测,本文所使用的特征融合方式和分层系数可能效果不是最佳的,但肯定是有效的。针对本文提出的算法模型提出一种具有普适性的融合方式和图像基本特征的选取是进一步的研究工作。

参 考 文 献

[1] WANG S. A Review of Gradient-Based and Edge-Based Feature Extraction Methods for Object Detection[C]// International Conference on Computer and Information Technology. IEEE, 2011:277-282.

[2] OLSON C F, HUTTENLOCHER D P. Automatic target recognition by matching oriented edge pixels[J]. IEEE Transactions on Image Processing, 1997, 6(1):103-113.

[3] GAVRILA D M, PHILOMIN V. Real-time object detection for smart vehicles[C]// The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999. IEEE, 1999: 87-93.

[4] CANNY J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986(6):679-698.

[5] HINTERSTOISSER S, LEPETIT V, ILIC S, et al. Dominant orientation templates for real-time detection of texture-less objects[C]// 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010:2257-2264.

[6] HINTERSTOISSER S, CAGNIART C, ILIC S, et al. Gradient response maps for real-time detection of textureless objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5):876-888.

[7] HSIAO E. Addressing ambiguity in object instance detection[D]. Pittsburgh: Carnegie Mellon University, 2013.

[8] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[9] KE Y, SUKTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2004:506-513.

[10] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded Up Robust Features[C]// European Conference on Computer Vision. Springer-Verlag, 2006:404-417.

[11] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling[C]// International Conference on Computer Vision. IEEE, 2010:32-39.

[12] LIENHART R, MAYDT J. An extended set of Haar-like features for rapid object detection[C]// International Conference on Image Processing. 2002:900-903.

[13] VIOLA P, JONES M. Fast and robust classification using asymmetric adaboost and a detector cascade[C]// Advances in Neural Information Processing Systems. 2002:1311-1318.

[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.

[15] SZEGEDY C, TOSHEV A, ERHAN D. Deep neural networks for object detection[C]// Advances in Neural Information Processing Systems. 2013:2553-2561.

[16] ERHAN D, SZEGEDY C, TOSHEV A, et al. Scalable object detection using deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:2147-2154.

[17] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv:1312.6229, 2013.

[18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:580-587.

[19] GIRSHICK R. Fast R-CNN[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:1440-1448.

[20] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.

[21] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1.

[22] MUNDER S, GAVRILA D M. An Experimental Study on Pedestrian Classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11):1863-1868.